Error Graphs and the Reconstruction of Elements in Groups

Vladimir I. Levenshtein*
Keldysh Institute of Applied Mathematics,
Russian Academy of Sciences, Moscow, Russia
leven@keldysh.ru

Johannes Siemons
School of Mathematics,
University of East Anglia, Norwich, UK
j.siemons@uea.ac.uk

Accepted Version, November 20, 2008; printed October 25, 2011

Abstract

Packing and covering problems for metric spaces, and graphs in particular, are of essential interest in combinatorics and coding theory. They are formulated in terms of metric balls of vertices. We consider a new problem in graph theory which is also based on the consideration of metric balls of vertices, but which is distinct from the traditional packing and covering problems. This problem is motivated by applications in information transmission when redundancy of messages is not sufficient for their exact reconstruction, and applications in computational biology when one wishes to restore an evolutionary process. It can be defined as the reconstruction, or identification, of an unknown vertex in a given graph from a minimal number of vertices (erroneous or distorted patterns) in a metric ball of a given radius r around the unknown vertex. For this problem it is required to find minimum restrictions for such a reconstruction to be possible and also to find efficient reconstruction algorithms under such minimal restrictions.

In this paper we define error graphs and investigate their basic properties. A particular class of error graphs occurs when the vertices of the graph are the elements of a group, and when the path metric is determined by a suitable set of group elements. These are the undirected Cayley graphs. Of particular interest is the transposition Cayley graph on the symmetric group which occurs in connection with the analysis of transpositional mutations in molecular biology [17, 19]. We obtain a complete solution of the above problems for the transposition Cayley graph on the symmetric group.

KEYWORDS: Reconstruction, Coding Theory, Biological Sequence Analysis, Cayley Graphs, Stirling Numbers

AMS CLASSIFICATION: 94A55, 94A15, 05E10, 05E30

*This research was supported by the Russian Foundation for Basic Research (Grant 04-01-00112).

1 Introduction: A Graph-Theoretical Approach to Efficient Reconstruction

The problem of the efficient reconstruction of sequences was introduced in [12, 13, 14] as a problem in coding theory, and similar questions about the efficient reconstruction of integer partitions were considered in [15, 18]. In this paper we discuss a graph-theoretical setting in which efficient reconstruction problems can be studied as a uniform theory.

Let $\Gamma = (V, E)$ be a simple, undirected and connected graph with vertex set V and edge set E. We regard the vertices in V as units of information in the given reconstruction problem, and for two vertices $x \neq y$ in V we regard $\{x, y\}$ as an edge of Γ if y is obtained from x, or vice versa x from y, by a *single error* or *single distortion* of information. We might say that x and y are erroneous single error representations of each other, and that Γ is a single error graph. The precise definitions can be found in Section 2. The task of the reconstruction problem now is to *restore* or *reconstruct* the original unit of information from sufficiently many erroneous representations of it. In other words, an unknown vertex x in Γ is to be identified by suitable knowledge about its neighbouring vertices in Γ.

We denote the path distance between two vertices x and y of Γ by $d(x, y)$ and we let $B_r(x) = \{ y \in V \ : \ d(x, y) \leq r \}$ be the ball of radius r centered at x. For given $r \geq 1$ denote by $N(\Gamma, r)$ the largest number N such that there exist a set $A \subseteq V$ of size N and two vertices $x \neq y$ with $A \subseteq B_r(x)$ and $A \subseteq B_r(y)$. Thus any $N+1$ distinct vertices are contained in $B_r(x)$ for at most one vertex x while there are some N vertices simultaneously contained in $B_r(x)$ and $B_r(y)$ for some $x \neq y$. This means that an unknown vertex of Γ can be identified, or reconstructed uniquely, by any set of $N(\Gamma, r) + 1$ or more distinct vertices at distance at most r from the vertex, provides that such a set exists.

In graph theoretical terms we are therefore required, for an arbitrary graph Γ and an integer $r \geq 1$, to determine the number

$$N(\Gamma, r) = \max_{x,y \in V,\ x \neq y} |B_r(x) \cap B_r(y)| \tag{1}$$

and to construct an efficient algorithm by which any unknown vertex x in V can be identified uniquely from an arbitrary set of $N(\Gamma, r)+1$ vertices at distance r or less from x. Evidently we can assume that r is at most $d(\Gamma)$, the diameter of Γ. Throughout the paper we assume that $d(\Gamma) \geq 2$ and in particular $|V| \geq 3$.

Problems of this kind have been solved for some graphs and metric spaces of interest in coding theory, and to give an impression of such results we review the example of Hamming spaces and Johnson spaces. The Hamming space F_q^n consists of q^n vectors of length n over the alphabet $\{0, 1, ..., q-1\}$ with metric $d(x, y)$ given by the number of coordinates in which the vectors x and y differ. This metric space can be represented by a graph Γ whose vertices are the vectors of F_q^n with two vectors connected by an edge if and only if they differ in a single coordinate. The path distance between two vertices then is the Hamming distance between the corresponding vectors. Therefore we can identify F_q^n with this graph Γ. In [12, 13, 14] it was shown that for any n, q and r we have

$$N(F_q^n, r) = q \sum_{i=0}^{r-1} \binom{n-1}{i} (q-1)^i \ . \tag{2}$$

Furthermore, any $x \in F_q^n$ can be reconstructed from $N = N(F_q^n, r)+1$ vectors of $B_r(x)$, written as the columns of a matrix, by applying the majority algorithm to the rows of the matrix.

For any $1 \leq w \leq n-1$ the Johnson space J_w^n consists of the $\binom{n}{w}$ binary vectors in F_2^n of length n and Hamming weight w, where distance is equal to half the (even) Hamming distance in F_2^n. This distance coincides with the minimal number of coordinate transpositions needed to transform one vector into the other. The Johnson space then can be viewed as a graph Γ whose vertices are the vectors of J_w^n with two vectors connected by an edge if and only if one is obtained from the other by a transposition of two coordinates. The path distance between two vertices of Γ then is the Johnson distance between the corresponding vectors. Therefore we can identify J_w^n with this graph Γ. In [12, 13] it was also shown that for any n, w and r we have

$$N(J_w^n, r) = n \sum_{i=0}^{r-1} \binom{w-1}{i} \binom{n-w-1}{i} \frac{1}{i+1} \; . \tag{3}$$

Furthermore, any $x \in J_w^n$ can be reconstructed from $N = N(J_w^n, r)+1$ vectors of $B_r(x)$, written as the columns of a matrix, by applying a threshold algorithm to the rows of the matrix.

In the first part of this paper we make the notion of error graphs precise and develop the theory needed to estimate $N(\Gamma, r)$ in some general situations. In this respect our main results are Theorems 1 and 2 which give lower bounds for $N(\Gamma, 1)$ and $N(\Gamma, 2)$ in terms of other graph parameters. It may be useful to mention that the idea of reconstructing a vertex in a given graph has nothing to do, a priory, with the classical Ulam problem of reconstructing a graph from the isomorphism classes of its vertex-deleted subgraphs. So we do not refer to the well-known and unresolved vertex-reconstruction problem. Nevertheless, error graphs are such a general tool that even this problem can be phrased suitably a problem on error graphs.

In the second part of the paper we deal with error graphs for which the vertex set consists of the elements of a group, and where the errors are defined by a certain set of group elements. Such graphs turn out to be undirected Cayley graphs, and in Sections 4 and 5 we show that many important error graphs occur as Cayley graphs. In Section 5 we discuss how transpositional errors in biological nucleotide sequences can be described as errors in the transposition Cayley graph $\mathrm{Sym}_n(T)$ on the symmetric group. The remainder of the paper deals with this graph in particular.

In Theorem 4 we determine the full automorphism group of the transposition Cayley graph $\mathrm{Sym}_n(T)$. The explicit value of $N(\mathrm{Sym}_n(T), r)$ can be found in Theorems 5, 6 and 7 for $1 \leq r \leq 3$. To state the main result on $N(\mathrm{Sym}_n(T), r)$ for arbitrary $r \geq 1$ let $c(n, n-r)$ be the number of permutations on $\{1..n\}$ having exactly $n-r$ cycles. Thus the $c(n, n-r)$ are the signless Stirling numbers of the first kind. We also need the following *restricted Stirling numbers*: Let $c_{3^1}(n, n-r)$ be the number of permutations g on $\{1..n\}$ having exactly $n-r$ cycles such that 1, 2 and 3 belong to the same cycle of g. The main result on $N(\mathrm{Sym}_n(T), r)$ is Theorem 9. It shows that for all $r \geq 1$

$$\begin{aligned} N(\mathrm{Sym}_n(T), r) &= \sum_{i=0}^{r-1} c(n, n-i) \\ &+ c_{3^1}(n, n-r) + c_{3^1}(n, n-(r+1)) \, . \end{aligned} \tag{4}$$

for all sufficiently large n. Furthermore, the maximum $N(\mathrm{Sym}_n(T), r) = |B_r(x) \cap B_r(y)|$ occurs for any $x \neq y$ for which $x^{-1}y$ is a 3-cycle on $\{1..n\}$. We mention the connection between this

theorem and the Poincaré polynomial of $\mathrm{Sym}_n(T)$. When Γ is an arbitrary finite graph and v a vertex of Γ let c_i denote the number of vertices at distance i from v. Then

$$\Pi_{\Gamma,v}(t) := \sum_{0\leq i} c_i t^i$$

is the *Poincaré polynomial* of Γ at v. When this polynomial is independent of v we write simply $\Pi_\Gamma(t)$. For the transposition Cayley graph $\mathrm{Sym}_n(T)$ the Poincaré polynomial is

$$\Pi_{\mathrm{Sym}_n(T)}(t) = \sum_{i=0}^{n-1} c(n, n-i)\, t^i \tag{5}$$

where the $c(n, n-i)$ are the Stirling numbers appearing in (4). This shows that the reconstruction parameters $N(\Gamma, r)$ are related to important graph invariants.

In this paper we have avoided technical terminology as far as possible in order to make this material accessible to non-specialists. For the same reasons we have added a few key references to texts in computing and computational biology.

2 Errors in Graphs

We will now fix the notation used for the remainder. Let $\Gamma = (V, E)$ be a finite graph with vertex set V and edge set E. All edges are undirected and there are no multiple edges or loops. Let x, y be vertices. Then x and y are *adjacent* to each other if $\{x, y\}$ is an edge. Further, $d(x, y)$ denotes the usual graph distance between the vertices, that is the length of a shortest path from x to y. Put $d(x, y) = \infty$ if x and y are in different components. For $i \geq 0$ we let $B_i(x) := \{y \in V \ : \ d(x, y) \leq i\}$ and $S_i(x) := \{y \in V \ : \ d(x, y) = i\}$ be the *ball* and *sphere* of radius i around x, respectively.

We put $k_i(x) = |S_i(x)|$ and for $y \in S_i(x)$ we set

$$\begin{aligned} c_i(x, y) &:= |\{z \in S_{i-1}(x) \ : \ d(z, y) = 1\}|, \\ a_i(x, y) &:= |\{z \in S_i(x) \ : \ d(z, y) = 1\}|, \\ b_i(x, y) &:= |\{z \in S_{i+1}(x) \ : \ d(z, y) = 1\}|. \end{aligned} \tag{6}$$

It is clear that $b_0(x, y) = k_1(x)$, that $a_1(x, y) = a_1(y, x)$ is the number of triangles over the vertices x and y, and that $c_2(x, y)$ is the number of common neighbours of x and $y \in S_2(x)$. Let

$$\begin{aligned} \lambda &= \lambda(\Gamma) = \max_{x,y\in V,\, d(x,y)=1} a_1(x, y) \\ \mu &= \mu(\Gamma) = \max_{x,y\in V,\, d(x,y)=2} c_2(x, y). \end{aligned} \tag{7}$$

Since $|B_r(x) \cap B_r(y)| > 0$ for $x \neq y$ only if $d(x, y) \leq 2r$ we have

$$N(\Gamma, r) = \max_{1\leq s\leq 2r} N_s(\Gamma, r) \tag{8}$$

where

$$N_s(\Gamma, r) = \max_{x,y \in V,\, d(x,y)=s} |B_r(x) \cap B_r(y)|. \tag{9}$$

In particular, $N_1(\Gamma, 1) = \lambda + 2$ and $N_2(\Gamma, 1) = \mu$ so that

$$N(\Gamma, 1) = \max(\lambda + 2, \mu). \tag{10}$$

Finding or estimating the value $N(\Gamma, r)$ for graphs of interest in applications is the main aim of our investigation here. We note the following general bounds for $N(\Gamma, r)$.

Lemma 1 *Suppose that $x \neq y$ are vertices in the connected graph $\Gamma = (V, E)$ at distance $s = d(x, y)$ from each other. Let $r \geq 0$ be an integer.*

(i) If $r \geq s$ then $B_r(x) \cap B_r(y) = B_{r-s}(x) \cup [(B_r(x) \setminus B_{r-s}(x)) \cap B_r(y)]$.In particular, we have $N_s(\Gamma, r) \geq |B_{r-s}(x)|$ and

$$N(\Gamma, r) \geq |B_{r-1}(x)| \tag{11}$$

for some $x \in V$.

(ii) If $r < s$ then $B_r(x) \cap B_r(y) = B_r(y) \cap [B_r(x) \setminus B_{s-r-1}(x)]$.

Proof: One should think of $B_r(x) \setminus B_{r-s}(x)$ as an annulus around x. (i) Starting on a path of length s from y to x any vertex in $B_{r-s}(x)$ can be reached by a further path of length at most $r - s$. The other statements are immediate from this. (ii) This is an direct consequence of the triangle inequality. □

We set

$$k_i(\Gamma) = \max_{x \in V} k_i(x) \quad \text{and} \quad k(\Gamma) = k_1(\Gamma). \tag{12}$$

Then Γ is *regular* of valency k (or *k-regular)* if all its vertices have constant valency $k = k(\Gamma)$. A k-regular graph is *distance-regular* if the numbers $c_i(x, y)$ and $b_i(x, y)$ (and hence $a_i(x, y) = k - c_i(x, y) - b_i(x, y)$) do not depend on $x \in V$ and $y \in S_i(x)$, for all $i = 0, 1, ..., d(\Gamma)$. A distance-regular graph of diameter 2 is *strongly regular.* A good reference to strongly regular graphs is Chapter 21 in [21] or also [4]. In such a graph there are integers λ and μ so that any pair of vertices $x \neq y$ is simultaneously adjacent to exactly λ vertices if $\{x, y\}$ is an edge, and to exactly μ vertices if $\{x, y\}$ is not an edge. Our use in (7) of the symbols λ and μ is therefore a natural extension to graphs which are not strongly regular.

Let $\mathrm{Aut}(\Gamma)$ be the automorphism group of Γ. If Γ is vertex-transitive (that is, for any two vertices in V there is an automorphism of Γ mapping one onto the other) then $k_i(x) = k_i(\Gamma)$ is constant for all $x \in V$ and i. In particular, such a graph is regular. However, even for vertex-transitive graphs the $c_i(x, y)$ and $b_i(x, y)$ usually depend on $y \in S_i(x)$, and this can cause difficulties in finding $N(\Gamma, r)$. This phenomenon can be observed already on relatively small graphs, see the Remark following Lemma 4.

The Hamming and Johnson graphs are examples of error graphs in which two vertices $x \neq y$ are joined by an edge if and only if there exists a single error (the substitution of a symbol or the transposition of two coordinates, respectively) which transforms x to y and there exists a single

error which transform y to x. This observation leads to a natural general theory of single errors which we began in [13]. For this we let V be a finite (or countable) set. A *single error* on V is an injection $h : V_h \to V$ defined on a non-empty subset $V_h \subseteq V$ so that $h(x) \neq x$ for all $x \in V_h$. A non-empty set H of single errors will be called a *single error set,* or just *error set,* provided the following two properties hold:

(i) For each $h \in H$ and $x \in V_h$ there exists some $g \in H$ so that $h(x) \in V_g$ and $g(h(x)) = x$, and

(ii) For all distinct pairs $x, y \in V$ there exist $x = x_1, x_2, ..., x_m = y \in V$ and $h_1, h_2, ..., h_{m-1} \in H$ such that $x_{i+1} = h_i(x_i)$, for $i = 1, ..., m-1$.

For such a set H we construct the *error graph* $\Gamma_H = (V, E)$ where $E = \{\, \{x, h(x)\} \ : \ x \in V \text{ and } h \in H \,\}$. Note, by the conditions on H we see that Γ_H has no loops and that all edges are undirected. The condition (ii) says that there is a path between any two vertices, and hence that Γ_H is connected. Furthermore, the usual path distance $d(x, y)$ on Γ_H now measures the minimum number of single errors required to transform x to y or y to x.

It is easily seen that every connected simple graph Γ can be represented as an error graph where we can assume in addition that the single error set consists of involutions (that is, partial maps h defined on suitable subsets of V such that $h^{-1} = h$). For if $c \colon E \to \{1...\chi\} \subseteq \mathbb{N}$ is an edge colouring of Γ then each fiber $c^{-1}(i)$ with $i = 1, ..., \chi$ defines a natural involutionary error h_i which is obtained by interchanging the two end vertices of any edge coloured by i. In particular, every connected graph Γ is an error graph with at most $\chi = \chi_E(\Gamma)$ errors where $\chi_E(\Gamma)$ is the edge-chromatic number of Γ.

By Vizing's theorem [22] this minimum number (over all Γ) is equal to $k(\Gamma) + 1$ where $k(\Gamma)$ is the maximum degree of Γ, as in (12). It is a natural question to ask whether any connected simple graph Γ can be represented as an error graph Γ_H for some error set H of cardinality $k(\Gamma)$. The answer is affirmative, see [13], where it is also shown that the property (i) can in general not be replaced by a stronger property $H = H^{-1}$ (meaning that $h^{-1} \in H$ if $h \in H$).

In the examples discussed before, the Hamming graph is an error graph when $V = F_q^n$ and when H consists of the $n(q-1)$ non-zero vectors $h \in F_q^n$ of Hamming weight 1, with action given by $h(x) = h + x$ for $x \in V$. Also the Johnson graph J_w^n is of this form when we view V as the set of all w-element subsets of $\{1, .., n\}$ and when H is the set of all $\binom{n}{2}$ transpositions (i, j) interchanging i and j in $\{1, .., n\}$, in their natural permutational action on V obtained by permuting the coordinates of vectors. In order to make sure that the single error property $h(x) \neq x$ holds for all vertices $x \in V_h$ one has to restrict the domain of (i, j) to those sets which contain exactly one of i and j.

Similarly, the insertion and deletion errors for finite sequences over an alphabet A can be described in this fashion as an infinite error graph. As vertex set we consider the set $V = A^0 \cup A^1 \cup A^2 \cup ... \cup A^n \cup ...$ of all finite words over A. As single error set we take $H := \{d_1, d_2, .., d_m, ...\} \cup \{i_1(a), i_2(a), ..., i_m(a), ... \ : a \in A\}$ where d_m deletes the m^{th} entry in any word of length $\geq m$ while $i_m(a)$ inserts a as the m^{th} entry in any sequence of length $\geq m-1$. As expected, the usual graph metric is indeed the Levenshtein error distance [11] for sequences. Situations where the model of undirected single error graphs is not applicable include asymmetric errors, some further comments can be found in [13].

3 Some Bounds for Regular Graphs

For the remainder we assume that Γ is a connected and regular graph on $v \geq 4$ vertices, with degree $2 \leq k = k(\Gamma)$ and parameters $\lambda = \lambda(\Gamma)$, $\mu = \mu(\Gamma)$. We have $0 \leq \lambda \leq k-1$, $1 \leq \mu \leq k$ and the diameter of Γ is $d(\Gamma) \geq 1$.

For some classes of regular and strongly regular graphs on v vertices we have $N(\Gamma, 1) = o(v)$ as $v \to \infty$. The following strongly regular graphs are well known, see Chapter 21 in [21] or [4]. The triangle graph $T(m)$ is strongly regular with parameters $v = m(m-1)/2$, $k = 2(m-2)$, $\lambda = m-2$, $\mu = 4$ and hence $N(T(m), 1) = m$. The lattice graph $L_2(m)$ is strongly regular with parameters $v = m^2$, $k = 2(m-1)$, $\lambda = m-2$, $\mu = 2$ and hence $N(L_2(m), 1) = m$. Meanwhile the Paley graphs $P(q)$ (q a prime congruent to 1 mod 4) is strongly regular with parameters $v = q$, $k = (q-1)/2$, $\lambda = (q-5)/4$, $\mu = (q-1)/4$ and hence $N(P(q), 1) = (q+3)/4$. The complement of a strongly regular graph Γ is also strongly regular (although not necessarily connected). This complementary graph $\overline{\Gamma}$ has parameters $v(\overline{\Gamma}) = v$, $k(\overline{\Gamma}) = v-k-1$, $\lambda(\overline{\Gamma}) = v-2k-2+\mu$, $\mu(\overline{\Gamma}) = v-2k+\lambda$, and hence $N(\overline{\Gamma}, 1) = v-2k+\max(\mu, \lambda)$.

Let $O_m^t = O_m \star O_m \star ... \star O_m$ be the product of t copies of the empty graph on m vertices. This is the complete t-partite graph with $v = tm$, each part consisting of m vertices and edges connecting vertices from different parts in all possible ways. If $t \geq 2$ this graph is connected and strongly regular.

The complete graph on v vertices is denoted by K_v. We recall that a 1-factor of a graph is a collection of disjoint edges covering all vertices (a complete matching of the vertices of Γ). When v is even consider the graph obtained from K_v by removing the edges of a 1-factor. This graph is strongly regular with parameters $k = \mu = v-2$, $\lambda = v-4$ and coincides with O_2^t with $t = \frac{v}{2}$. Conversely, if Γ is a regular of degree $k = v-2$ then v is even and $\Gamma = O_2^t$ with $t = \frac{v}{2}$. When $t = \frac{v}{2}$ then $N(O_2^t, 1) = \lambda + 2 = v - 2 = \frac{1}{2}(v+\lambda)$. More generally we have:

Theorem 1 *Let Γ be a regular graph with $k \leq v-2$. Then we have*

$$N(\Gamma, 1) \leq \frac{1}{2}(v+\lambda) \tag{13}$$

with equality if and only if $k-\lambda = v-k$ divides v and Γ is the strongly regular graph O_m^t with $m = k-\lambda$ and $v = tm$.

Proof: By (10) we have $N(\Gamma, 1) = \max\{\lambda+2, \mu\}$. If $\Gamma = O_m^t$ with $v = tm$ then $k = v-m$, $\mu = v-m$ and $\lambda = v-2m$ so that $N(\Gamma, 1) = \mu = \frac{1}{2}(v+\lambda)$. For the converse assume first that $\lambda = k-1$. In this case (10) implies that $N(\Gamma, 1) = k+1$ which is not possible as $k+1$ is the cardinality of any single ball. Therefore $\lambda \leq k-2$ and from the assumptions in the theorem it follows that $\lambda \leq v-4$ or $\frac{1}{2}\lambda + 2 \leq \frac{1}{2}v$. Hence $\lambda + 2 \leq \frac{1}{2}(v+\lambda)$ with equality if and only if $\lambda = v-4$. In the latter case only $k = v-2$ is possible and so we have the situation already discussed, Γ is O_m^t with $m = k-\lambda = v-k = 2$ and $v = 2t$.

It is left to show that $\mu \leq \frac{1}{2}(v+\lambda)$ and to find the conditions for equality. For a k-regular graph (V, E) and a vertex x in V we count the number of edges between $S_1(x)$ and $S_2(x)$. This gives

$$\sum_{y \in S_1(x)} (k - 1 - a_1(x,y)) = \sum_{z \in S_2(x)} c_2(x,z),$$

see again the definitions in (7). This gives $k(k-1-\lambda) \le \mu k_2(x)$ and since $k_2(x) \le v-k-1$ we obtain

$$k(k-1-\lambda) \le \mu k_2(x) \le \mu(v-k-1). \tag{14}$$

Since $1 \le \mu \le k$ we get $k-1-\lambda \le v-k-1$ and hence $\mu \le k \le \frac{1}{2}(v+\lambda)$ as required.

If $\mu = k = \frac{1}{2}(v+\lambda)$ then we have equalities in (14). For a regular graph it is well-known that the inequalities in (14) turn into equalities if and only if the graph is strongly regular, see for instance PROBLEM 21A in [21]. So let Γ be strongly regular with $\mu = k$. Then any pair of distinct and non-adjacent vertices have the same k neighbours. It follows that $x = x'$ or x is not adjacent to x' defines an equivalence relation on the vertices of Γ, with all equivalence classes of size $m := v-k$. Hence m divides $v = tm$ and $\Gamma = O_m^t$. □

Theorem 2 (Linear Programming Bound) *Let Γ be a regular graph of valency $k \ge 2$. Then*

$$N_2(\Gamma,2) \ge \mu\left(k-1-\frac{1}{2}(\mu-1)(N(\Gamma,1)-2)\right)+2. \tag{15}$$

We note that this rather general bound is quite sharp, see the comment following Theorem 6.

Proof: There are two vertices $x, x' \in V$ with $d(x,x') = 2$ so that the set $Y = \{y_1, .., y_\mu\}$ of all vertices at distance 1 from both x and x' has $\mu \ge 1$ elements. If $\mu = 1$ then y_1 has $k-2$ neighbours other than x and x'. It follows that $N_2(\Gamma,2) \ge |B_2(x) \cap B_2(x')| \ge 3+k-2$ and so (15) holds. Hence we assume that $\mu \ge 2$.

Let $U = \bigcup_{i=1}^{\mu} B_1(y_i) \backslash \{x, x'\}$. We show that the number of elements in U is at least $\mu\left(k-1-\frac{1}{2}(\mu-1)(N(\Gamma,1)-2)\right)$. For $h = 1, ..., \mu$ let $U(h)$ be the vertices of U which belong to exactly h of the sets $B_1(y_i)$, as $i = 1, ..., \mu$. In particular, $U = U(1) \cup ... \cup U(\mu)$ is a partition and so

$$|U| = \sum_{h=1}^{\mu} |U(h)| \ .$$

Next observe that the set $\{(z,\, B_1(y)) \,:\, y \in Y \text{ and } z \in B_1(y) \cap U\}$ has cardinality

$$\sum_{h=1}^{\mu} h|U(h)| = \mu(k-1)$$

and the set $\{(z,\, \{B_1(y),\, B_1(y')\}) \,:\, y \ne y' \in Y \text{ and } z \in B_1(y) \cap B_1(y') \cap U\}$ has cardinality

$$\sum_{h=2}^{\mu} \binom{h}{2} |U(h)| = \sum_{\{y,y'\} \subseteq Y,\, y \ne y'} (|B_1(y) \cap B_1(y')| - 2) \le \binom{\mu}{2} (N(\Gamma,1)-2) \ .$$

The last inequality holds as $y \ne y'$ implies $|B_1(y) \cap B_1(y')| \le N(\Gamma,1)$.

Set $u_h := |U(h)|$ for $h = 1, ..., \mu$ and $u := |U|$. To find a lower bound for u we minimize

$$u - \mu(k-1) = -1u_2 - 2u_3 - ... - (\mu-1)u_\mu$$

for the non-negative integers $u_2, ..., u_\mu$ subject to the constraints

$$\sum_{h=2}^{\mu} hu_h \leq \mu(k-1)$$

and

$$\sum_{h=2}^{\mu} \binom{h}{2} u_h \leq \binom{\mu}{2} (N(\Gamma,1)-2) \ .$$

Let $u^* \leq u - \mu(k-1)$ be the required minimum. Then by the duality of linear programming, see for instance Section 7.5 in [16], the value of u^* maximizes

$$-\mu(k-1)n_1 \ - \ \binom{\mu}{2} (N(\Gamma,1)-2)\, n_2$$

subject to $n_1, n_2 \geq 0$ and the dual constraints

$$hn_1 + \binom{h}{2} n_2 \geq h-1 \quad \text{for} \quad h = 2, ..., \mu \ .$$

Note that $n_1 = 0$ and $n_2 = 1$ satisfies the dual constraints for all $\mu \geq 2$ and hence

$$u - \mu(k-1) \geq u^* \geq -\binom{\mu}{2} (N(\Gamma,1)-2) \ .$$

Therefore $u \geq \mu\left(k - 1 - \frac{1}{2}(\mu-1)(N(\Gamma,1)-2)\right)$ as required. □

Note for instance that $N_2(\Gamma,2) \geq k+1$ when $\mu = 1$, $N_2(\Gamma,2) \geq 2k$ when $\mu = 2$ and $N(\Gamma,1) = 2$, and $N_2(\Gamma,2) \geq 3k-4$ when $\mu = 3$ and $N(\Gamma,1) = 3$.

Corollary 1 *Suppose that Γ is a regular graph of valency k with no triangle nor pentagons. If $\mu \geq 2$ and $k \geq 1 + \binom{\mu}{2}$ then*

$$N_2(\Gamma,2) \geq N_1(\Gamma,2).$$

Proof: We have $\lambda = 0$ since Γ has no triangles so that $N(\Gamma,1) = \mu$ by (10). Similarly, $N_1(\Gamma,2) = 2k$ as Γ contains no pentagons. Using (15) we get

$$\begin{aligned} N_2(\Gamma,2) - 2k &\geq \mu\left(k - 1 - \frac{1}{2}(\mu-1)(N(\Gamma,1)-2)\right) + 2 - 2k \\ &= \mu\left(k - 1 - \frac{1}{2}(\mu-1)(\mu-2)\right) + 2 - 2k \\ &= (\mu-2)\left(k - 1 - \binom{\mu}{2}\right) \geq 0 \end{aligned}$$

and this completes the proof. □

4 Single error sets as group generators

An important class of graphs associated to single error sets is obtained when the vertex set of the graph are the elements of a finite group. So we let G be a finite group and consider the elements of $G = V$ as the vertices of the error graph $\Gamma = \Gamma_H$ for some error set H. The neutral element of G is denoted by $e = e_G$ and $1 = \{e_G\}$ is the identity subgroup of G. We suppose that the single error set *is determined as a subset* H *of* G so that the action of errors on vertices is given by the group product. That is, if $h \in H$ and $x \in G$ then $h(x) := xh^{-1}$. In this situation H is a single error set if and only if H does not contain e_G and

(i) H satisfies $H = H^{-1}(= \{\, h^{-1} \;:\; h \in H \,\})$, and

(ii) H generates G as a group.

The first condition is clear since there is some g in H with $g(h(x)) = (xh^{-1})g^{-1}$
$= x$ for a vertex x in V if and only if $g = h^{-1}$ belongs to H. The second condition is a restatement of the connectedness of the error graph. Note that we have set $h(x) := xh^{-1}$, rather than $h(x) := xh$. This is advisable so that the multiplication of errors as elements of G agrees with the co-cattenation of the corresponding maps, $(gh)(x) = x(gh)^{-1} = xh^{-1}g^{-1} = g\big(h(x)\big)$.

As is well known, in this situation Γ_H is the *undirected Cayley graph* on G for the generating set H, and H is the *Cayley set* for Γ_H. Note conversely that every undirected Cayley graph can be viewed as a single error graph.

In the following we review some of the theory of Cayley graphs from the viewpoint of single error graphs. Let H be a Cayley set in the finite group G with corresponding graph $\Gamma_H = (V, E)$ on the vertex set $V = G$ and let $\mathrm{Aut}(\Gamma_H)$ be the automorphism group of Γ_H. We consider two basic kinds of automorphisms of Γ_H. For each g in G the left-multiplication on V, with $g\colon x \mapsto gx$ for $x \in V$, induces an automorphism of Γ_H since $g\colon \{x, xh^{-1}\} \mapsto \{gx, gxh^{-1}\}$ maps edges to edges. If we think of $\{x, xh^{-1}\}$ as being labelled by $\overline{h} = \{x^{-1}(xh^{-1}), (hx^{-1})x\} = \{h^{-1}, h\}$, the quotients of its end vertices, then $\{gx, gxh^{-1}\}$ has the same label as $\{x, xh^{-1}\}$. Therefore left-multiplication by elements of G are automorphisms that preserves all edge labels.

This action is transitive on vertices and only the identity element fixes any vertex. This is therefore the *regular action* of G on itself. This property characterizes Cayley graphs: Γ is the Cayley graph of some group if and only if Γ admits a group of automorphisms that acts regularly on its vertices, see for instance Chapter 6 in [1]. Note however that the graph usually does not determine the group.

We now describe graph automorphisms that change edge labels. Let C be a group of automorphisms of G as a group. For the action of $\beta \in C$ we write $\beta\colon x \mapsto \beta(x)$ and so $\beta(xy) = \beta(x)\beta(y)$ as β is an automorphism of the group structure. We will also require that C preserves H, in the sense that $\beta(h) \in H$ for all $h \in H$ and $\beta \in C$. Then C is a group of automorphisms of Γ_H since $\beta\colon \{x, xh^{-1}\} \mapsto \{\beta(x), \beta(xh^{-1})\} = \{\beta(x), \beta(x)\beta(h)^{-1}\}$ maps edges to edges, as $\beta(h)^{-1} = \beta(h^{-1}) \in H$. Note that the label of $\beta(\{x, xh^{-1}\})$ now is $\overline{\beta(h)}$.

The semi-direct product $G \cdot C$ is the (abstract) group of all pairs (g, β) with multiplication $(g', \beta')(g, \beta) = (g'\beta'(g), \beta'\beta)$. It acts on the graph as automorphisms by setting

$$(g, \beta)\colon x \mapsto g\beta(x) \quad \text{for } x \in V.$$

This gives an injective group homomorphism from $G{\cdot}C$ to $\mathrm{Aut}\,\Gamma_H$ so that we can regard $G{\cdot}C$ as a subgroup of $\mathrm{Aut}\,\Gamma_H$. We collect these facts:

Proposition 1 *Let Γ_H be the error graph on the group $V = G$ with error set H. Then the left-multiplication of vertices by elements of G forms a group of automorphisms of Γ_H which acts regularly on the vertex set V. If C is a group of automorphisms of G (as a group) such that $\beta(H) \subseteq H$ for all β in C then the semi-direct product $G{\cdot}C$ is contained in the automorphism group of Γ_H.*

A common example of this situation occurs when we consider conjugation by group elements. Let $b \in G$. Then *conjugation by b* is the automorphisms

$$x \mapsto bxb^{-1} =: x^b \quad \text{for } x \in G$$

and

$$x^G := \{x^b \ : \ b \in G\}$$

is the *conjugacy class* of x. In this case the error set H is invariant under conjugation if and only if H is a union of conjugacy classes. For C we then take the group $C := G/Z$ where $Z = Z(G) = \{b \in G \ : \ x^b = x \text{ for all } x \text{ in } G\}$ is the *center* of G. These are the *inner automorphisms* of G. So in this situation $G{\cdot}C$ is a group of automorphism of Γ_H. In Chapter 5 we will analyse this example further when G is the symmetric group Sym_n on the set $\{1..n\}$ and when H is the set of all transpositions on $\{1..n\}$. There we shall see that the full automorphism group of the error graph can be larger than $G{\cdot}C$, even if C is the group of all automorphisms of G as a group.

Another interesting example occurs when Γ is the Hamming graph. Here G is the vector space F_q^n where F_q is the field of q elements and H is the set of all vectors of the shape $(0, .., 0, a, 0, .., 0)$ with $a \neq 0$. Then G acts on itself as a group of translations, that is, maps of the kind $g\colon x \mapsto g{+}x$ for all $x \in F_q^n$. For C we can take the *monomial subgroup* $C = (F_q^\times)^n \cdot \mathrm{Sym}_n \subseteq \mathrm{GL}(n, q)$ acting naturally as linear maps on V. More precisely, C is the group of all $n \times n$ matrices with exactly one element from the multiplicative group $F_q^\times$ in each row and column. So here $F_q^n \cdot C$ is a group of affine linear maps on F_q^n that acts naturally as automorphisms on the Hamming graph Γ.

Considering again the general case we let $\Gamma_H = (G,\, E)$ be an error graph with error set H. We have seen that any group automorphism β fixing H as a set induces an automorphism of Γ_H. Evidently β also fixes the identity element $e = e_G$ in G. Assume therefore more generally that C is a group of automorphisms of Γ_H which fixes e. For any $x \in V$

$$x^C := \{\beta(x) \ : \ \beta \in C\}$$

is the *orbit* of x under C. In order to analyze the parameters $k_i(x),\ a_i(x, y),\ b_i(x, y)$ and $c_i(x, y)$ note that Γ_H is vertex transitive and therefore it suffices to consider the spheres with center e_G. Hence we abbreviate all parameters, writing $S_i,\ B_i,\ k_i = |S_i|,\ a_i(y),\ b_i(y)$ and $c_i(y)$, suppressing the reference to $x = e_G$ in each case. In general these parameters still depend on y although automorphisms provide at least for some form of regularity:

Proposition 2 *Let Γ_H be the error graph on the group $V = G$ with error set H and suppose that C is a group of automorphisms of Γ_H which fixes $e = e_G$. Then for each $i \geq 0$ the sphere $S_i = S_i(e)$ is a union of C-orbits.*

Further, suppose that y *and* y' *belong to the same* C*-orbit and that* $r, t \geq 0$*. Then* $|S_r \cap S_t(y)| = |S_r \cap S_t(y')|$ *and* $|B_r \cap B_t(y)| = |B_r \cap B_t(y')|$*. In particular,* $a_i(y) = a_i(y')$*,* $b_i(y) = b_i(y')$ *and* $c_i(y) = c_i(y')$ *for all* $i \geq 0$*.*

Proof: Let $y \in S_i$ and let $e, y_1, ..., y_i = y$ be a shortest path from e to y. If $\beta \in C$ then it is clear that $\beta(e) = e, \beta(y_1), ..., \beta(y_i) = \beta(y)$ is a shortest path from e to $\beta(y)$. It follows that $\beta(S_i) = S_i$ is a union of C-orbits. Now suppose that $y' = \beta(y)$. Then $\beta(S_i(y)) = S_i(\beta(y)) = S_i(y')$ and so $|S_r \cap S_t(y)| = |\beta(S_r \cap S_t(y))| = |\beta(S_r) \cap \beta(S_t(y))| = |S_r \cap S_t(y')|$. The remainder follows immediately, including the statement on a_i, b_i and c_i since these numbers are of the shape $|S_r \cap S_t(y)|$ for particular choices of r and t. □

If H is the single error set of Γ_H we set $H^0 := \{e_G\}$ and $H^i := HH^{i-1}$ inductively for $i > 0$. Clearly, Γ_H is regular of degree $k(\Gamma) = |H|$. If as before S_i denotes the sphere of radius i around $e = e_G$ then evidently $S_1 = H^1 = H$, $S_2 = H^2 \setminus (H^1 \cup H^0)$ and more generally,

$$S_i = H^i \setminus (H^{i-1} \cup H^{i-2} \cup ... \cup H^1 \cup H^0).$$

The following is easily shown and gives the value of $N(\Gamma_H, 1)$ by using (10).

Lemma 2 *In an error graph* Γ_H *with error set* H *we have*

$$\lambda(\Gamma_H) = \max_{x \in S_1} |\{(h, h') : x = hh' \text{ with } h, h' \in H\}| \quad \text{and}$$

$$\mu(\Gamma_H) = \max_{x \in S_2} |\{(h, h') : x = hh' \text{ with } h, h' \in H\}| \ .$$

5 Permutations distorted by transpositional errors

In the following we consider Cayley graphs when $G = \mathrm{Sym}_n$ is the symmetric group acting on the set $\{1..n\}$. Any subset H of G which generates G with $e \notin H$ and $H = H^{-1}$ is a Cayley set for G.

We express permutations in the usual cycle notation. (Throughout the word 'cycle' always refers to a particular kind of permutation, and never to a graph or subgraph.) A *transposition* on $\{1..n\}$ is a permutation of the shape $x = (i, j)$ with $1 \leq i \neq j \leq n$ if we suppress the 1-cycles of x. Particularly important graphs occur when $H = \{(1, 2), (2, 3), ..., (n-1, n)\}$ are the $n-1$ *Coxeter generators* of the symmetric group. These form a minimal set of transpositions needed to generate Sym_n. This set corresponds to the fundamental reflections associated to a chamber for the A-type Dynkin diagram. The chambers give rise to a triangulation of the euclidean unit sphere in $\mathbb{R}^{n-1}$. In this situation the graph distance function $d(x, y)$ in Γ_H is a discretized version of the geodetic distance on this sphere and presents the distance between two facets in the triangulation of the sphere, see for instance the book [2] of Grove and Benson on finite reflection groups. In this interpretation $B_r(x)$ is the 'cap' of facets on the sphere at distance $\leq r$ from the facet x and $N(\Gamma, r)$ is the number facets common to two such caps, with suitable distinct centers. Note also that here $d(e, -)$ evaluated for a single variable is the word length function in the corresponding Weyl group. This Cayley graph is of considerable importance in Lie theory and in many other parts of mathematics and physics. For a recent treatment of its combinatorics

we refer to [3]. We add that in computer science this graph is known as the *bubble–sort Cayley graph* and is used as a model for interconnection networks [8, 9]. Various other Cayley graphs on Sym_n have been considered in the literature, we mention in particular Diaconis' book [5] where metrics on Sym_n more generally are discussed.

By contrast we may consider the error graph on Sym_n when the single error set H consists of *all* transpositions (i, j) on $\{1..n\}$. This clearly is a highly redundant system of generators, situated at the other extreme to the case of the Coxeter elements in Sym_n which form a minimal generating set. In this situation a single error (i, j) transforms the vertex x to its neighbour $x(i, j)$ and all choices for $1 \leq i \neq j \leq n$ are admissible. A graph Γ_H of this type will be called a *transposition Cayley graph,* and these graphs are the subject of the remainder of the paper.

It may be useful to describe errors of this kind in a slightly more general setting. Let A be a finite alphabet with $|A| \geq 2$ and let A^n be the set of all words of length n over A. Then the single transposition error (i, j) on the coordinates of A^n is the map $(i, j) : a = (a_1, ..., a_i, ..., a_j, ..., a_n) \mapsto a^{(i,j)} = (a_1, ..., a_j, ..., a_i, ..., a_n)$ with all other entries of a unchanged. This gives rise to an error distance d_A on A^n where $d_A(a, b)$ is the least number of single transposition errors needed to transform a to b, if this is possible. In this case we must have $b = a^g$ for some $g \in \mathrm{Sym}_n$ and $d_A(a, b) \leq d(e_G, g)$ where the latter denotes the distance in the transposition Cayley graph. (Observe that $d_A(a, a^{(i,j)}) = 0$ if and only if $a_i = a_j$ while $d\,(e_G, (i, j)) = 1$ independently.) Note that this distance function defines a graph on A^n. Each component is an error graph with involutory errors (i, j) if we restrict the domain of the single error (i, j) to the words a in which $a_i \neq a_j$. In this way the transposition error graph Γ_H can be said to control the transposition errors on A^n.

In molecular biology transpositional errors are one of the three known mechanisms in the mutation and evolution of genetic information. The so-called *replication slippage* applied to a nucleotide sequence is a process that results in some strings of consecutive nucleotides being reversed or repeated in the sequence. Such replication slippages usually recur and give rise to so-called microsatellites which contain a high degree of information about the evolutionary process undergone by the nucleotide sequence in question, and often this happens in the non-coding part of the nucleotide sequence. For general information see Futuyma's book [7] on evolutionary biology as well as [17] and [19].

Replication slippage is therefore a combinations of two kinds of errors on sequences, on the one hand the insertion-deletion process already mentioned at the end of Section 2 and the transpositional errors in the transposition Cayley graph on the other. It may be worth to mention that the other principal mutation mechanisms are *point mutations* referring to the replacement of one nucleotide by another, and *frame shifts* which are the insertion or deletion of a group of nucleotides. Both of these are therefore covered by the insertion-deletion process.

Evidently any interval transposition or reversal (of a part of a nucleotide sequence) can be expressed as a product of single transpositional errors. However, it should be interesting to introduce such products as new single errors, and to consider the resulting error graph on Sym_n. A second point of interest should be to study the *resistance to transpositional errors:* As the nucleotide alphabet consists of just four letters, a single transpositional error is expressed only in a small proportion of all possible words in A^n, leaving many others unchanged by that error.

Returning to the general discussion of the transposition Cayley graph we note the following conventions. Permutations in Sym_n are multiplied from the right so that $(xy)(j) = x(y(j))$ for

all x, $y \in \mathrm{Sym}_n$ and $j \in \{1..n\}$. If x is written as a product of h_i disjoint cycles of length i for $1 \le i \le n$ then the *cycle type* of x is denoted as $\mathrm{ct}(x) = 1^{h_1} 2^{h_2} ... n^{h_n}$. Here it is essential to include 1-cycles so that $\sum_i i\, h_i = n$. As is well-known, two permutations are conjugate to each other through an element of Sym_n if and only if they have the same cycle type. Writing $G = \mathrm{Sym}_n$ therefore the conjugacy class

$$(1^{h_1} 2^{h_2} ... n^{h_n})^G := x^G = \{\, g^{-1} x g \;:\; g \in G \,\}$$

is the set of all permutations having the same cycle type as x.

We let $H = T := \{\, (i,j) \in \mathrm{Sym}_n \;:\; 1 \le i \ne j \le n\} = (1^{n-2}\, 2^1)^G$ be the set of all transpositions of $\{1..n\}$. Thus Γ_T is the transposition Cayley graph on Sym_n and will be denoted by $\mathrm{Sym}_n(T)$. The following collects some easily established facts.

Lemma 3 *For $n \ge 3$ the transposition Cayley graph $\mathrm{Sym}_n(T)$ is a connected $\binom{n}{2}$-regular graph of order $n!$ and diameter $n-1$. It is t-partite for any $2 \le t \le n$.*

Proof: The group Sym_n has order $n!$ and is generated by its $\binom{n}{2}$ transpositions. Its diameter is at most $n-1$ since any permutation is a product of at most $n-1$ transpositions. On the other hand, an n-cycle can not be written in terms of fewer than $n-1$ transpositions. No two elements in the same sphere S_i could be adjacent to each other as they have the same determinant $(-1)^i$. Hence $S_0,\ S_1,\ S_2,\ ...,\ S_{n-1}$ is a partition into n parts from which a t-partition can be obtained for any $t \le n$. □

For the product of a permutation with a transposition the following simple rule is essential. If $x = (i_1, .., i_k)(j_1, .., j_\ell)$ consists of two disjoint cycles and if $t = (i,j)$ interchanges elements from different cycles, say $i_1 = i$ and $j_1 = j$ without loss of generality as the cycles are determined only up to cyclic reordering, then

$$xt = (i_1, j_2, j_3, .., j_\ell, j_1, i_2, i_3, .., i_k) =: s \tag{16}$$

is a single cycle obtained by joining up the two cycles of x. Conversely, upon multiplying this equation again by t, we see that multiplying the single cycle s by a transposition of some two elements from that cycle gives $x = xtt = st$, hence splitting that single cycle into two cycles. Therefore multiplying any permutation x by a transposition results in a permutation which either joins up two cycles of x or splits one cycle of x into two, with no other changes.

Following the earlier convention whereby $S_i = S_i(e)$ we have that $H = T = S_1$ consists of all transpositions, S_2 consists of all 3-cycles (i,j,k) and all *double transpositions* $(i,j)(k,\ell)$ with i, j, k, ℓ distinct, and so on. As multiplication by a transposition increases or decreases the number of cycles by one it follows by induction that S_i consists of all permutations expressible as a product of $n-i$ disjoint cycles, counting also all 1-cycles.

The path distance between two permutations x and y is the least number d of transpositions t_i such that $xt_1...t_d = y$. Equivalently d is the least number of transpositions needed to write $x^{-1}y$ and also equal to the number of bisections and gluings needed to transform the cycles of x into those of y. The number of distinct paths from x to y is equal to the number of paths from e to $x^{-1}y$ and about these the following theorem gives complete information. It is based on Ore's theorem on the number of trees with n labeled vertices, see also Theorem 2 in [10].

Theorem 3 *[6] Suppose that x has cycle type $\mathrm{ct}(x) = 1^{h_1}2^{h_2}...n^{h_n}$ and let $1 \le i \le n-1$ be such that $\sum_{j=1}^{n} h_j = n - i$ is the number of cycles in x. Then the number of distinct ways to express x as a product of i transpositions is equal to*

$$i! \prod_{j=1}^{n} \left(\frac{j^{j-2}}{(j-1)!} \right)^{h_j} . \tag{17}$$

By the discussion above x cannot be written in fewer than i transpositions. The special case $i = n-1$ and $h_n = 1$ means that each of the $(n-1)!$ cycles of length n has n^{n-2} different representations as a product of $n-1$ transpositions. This number coincides with the number of trees with n labelled vertices, see also Section 5.3 in Stanley [20].

Let $1 \le i \le n-1$. If $y \in S_i$ has cycle type $\mathrm{ct}(y) = 1^{h_1}2^{h_2}..n^{h_n}$ and consists of $\sum_{j=1}^{n} h_j = n-i$ cycles then y is a product of i transpositions. As $\det y = (-1)^i$ we must have $a_i(y) = 0$. As a single cycle of length j can be split into two cycles as in (16) in $\binom{j}{2}$ different ways, we have $c_i(y) = \sum_{j=1}^{n} \binom{j}{2} h_j$. From $\sum_{j=1}^{n} j h_j = n$ it follows that $c_i(y) = \sum_{j=1}^{n} \binom{j}{2} h_j = \frac{1}{2}\left(\sum_{j=1}^{n} j^2 h_j - n \right)$. If we regard y as an element of Sym_m with $m > n$ then it is clear from (16) that $c_i(y)$ is independent of n. Finally, $b_i(y) = \binom{n}{2} - c_i(y)$. We collect these facts:

Lemma 4 *In $\mathrm{Sym}_n(T)$ the set S_i, where $1 \le i \le n-1$, consists of all permutations of $\{1..n\}$ which are composed of exactly $n-i$ disjoint cycles, including 1-cycles.*

If $y \in S_i$ has cycle type $\mathrm{ct}(y) = 1^{h_1}2^{h_2}..n^{h_n}$ then

$$c_i(y) = \frac{1}{2}\left(\sum_{j=1}^{n} j^2 h_j - n \right),$$

$$a_i(y) = 0 \quad \textit{and}$$

$$b_i(y) = \frac{1}{2}\left(n^2 - \sum_{j=1}^{n} j^2 h_j \right).$$

If y is regarded as an element in Sym_m with $m > n$ then only $b_i(y)$ depends on n.

REMARK: Loosely speaking, if y belongs to S_i we can think of $c_i(y)$ as the 'downward' degree of y, namely the number of neighbours of y in the next lower sphere S_{i-1}. The fact that this degree is independent of n will be used later on. Similarly $\binom{n}{2} - c_i(y)$ is the 'upward' degree of y. The transposition Cayley graphs are not distance-regular and they illustrate the fact that the up- and downward degrees are not constant for elements in the same sphere. This can be seen already in $\mathrm{Sym}_4(T)$. If y in S_2 is a 3-cycle then $c_2(y) = 3$ according to the three choice of a transposition splitting the 3-cycle. On the other hand, if $y = (1,2)(3,4)$ in S_2 is a double transposition then $c_2(y) = 2$ as there are just two ways to split one of the two cycles. This is true for any $n \ge 4$.

Next we discuss the automorphism group of the transposition Cayley graph. As before let $G = \mathrm{Sym}_n = V$ and set $\Gamma = \mathrm{Sym}_n(T)$. Let (a,b) be an element of the direct product $G \times G$. Then $(a,b) : x \mapsto axb^{-1}$ for $x \in V$ is an automorphism of Γ since for any transposition t we have

$xt \mapsto axtb^{-1} = (axb^{-1})(btb^{-1})$ in which btb^{-1} again is a transposition. Note that only the identity of $G \times G$ fixes all vertices since $axb^{-1} = x$ for all $x \in V$ implies that $a = b$ and hence that $a \in Z(G) = 1$. This implies that we can view $G \times G$ as a subgroup of $\mathrm{Aut}(\Gamma)$. Recall the discussion in Section 4. If we let C be the group of conjugation automorphisms, $x \mapsto x^b$ for all $b \in G$, then C is the *diagonal subgroup* $\{(b,b) : b \in G\} \subseteq G \times G$. Furthermore, we have $G \times G = G \cdot C$ as subgroups of $\mathrm{Aut}(\Gamma)$.

A further automorphism of Γ comes from the inversion map

$$\imath \colon x \leftrightarrow x^{-1} \quad \text{for } x \in V.$$

While $\imath$ is not an automorphism of the group it is an automorphism of the graph. For if $\{x, y\}$ is an edge with $y = xt$ and t a transposition then $y^{-1} = x^{-1}(yty^{-1})$ where yty^{-1} is a transposition and so $\{y^{-1}, x^{-1}\}$ is an edge. Since $(\imath(a,b)\imath)(x) = (ax^{-1}b^{-1})^{-1} = (b,a)(x)$ for all $x \in V$ we see that $\imath$ normalizes $G \times G$ by interchanging the two direct factors. This shows that the semi-direct product $(\mathrm{Sym}_n \times \mathrm{Sym}_n) \cdot \langle \imath \rangle$ is contained in $\mathrm{Aut}(\mathrm{Sym}_n(T))$.

Theorem 4 *For $n \geq 3$ the full automorphism group of $\mathrm{Sym}_n(T)$ is the semi-direct product $(\mathrm{Sym}_n \times \mathrm{Sym}_n) \cdot C_2$ where $C_2 = \langle \imath \rangle$ is the group of order 2 obtained by inverting the elements in $V = \mathrm{Sym}_n$.*

Proof: As before set $G = \mathrm{Sym}_n$, $\Gamma = \mathrm{Sym}_n(T)$ and let A be the group of all automorphisms of Γ. When $n = 3$ when Γ is the complete bipartite graph $K_{3,3}$ and in this case the statement can be checked directly from the description of the action of $(\mathrm{Sym}_3 \times \mathrm{Sym}_3) \cdot \langle \imath \rangle$ on Γ.

Now suppose that $n > 3$ and let $\alpha'' \in A$. As G acts vertex transitively by left-multiplication we select $g'' \in (G \times 1) \subseteq (G \times G)$ such that $\alpha' := g''\alpha''$ fixes e_G. This implies that α' fixes S_r as a set, for all $r \geq 1$, see Proposition 2. Furthermore, α' fixes each of the two conjugacy classes $(1^{n-3}3^1)^G$ and $(1^{n-4}2^2)^G$ in S_2 since $c_2 = 3$ on the first class while $c_2 = 2$ on the second class, see the remark following Lemma 4.

For $1 \leq i \leq n$ let the *pencil* P_i be the set $P_i = \{(i,j) \in S_1 : 1 \leq j \leq n \quad \text{and} \quad i \neq j\}$. Then the following holds: any pair $x \neq y \in P_i$ has exactly two joint neighbours in $(1^{n-3}3^1)^G$, and P_i is a maximal subset of S_1 with this property. Conversely, any set of $n-1$ vertices in S_1 satisfying this property is a pencil. Since $(1^{n-3}3^1)^G$ is invariant under α' we see that $\alpha'(P_i)$ again is a pencil. On the other hand, the diagonal element $(g,g) \in G \times G$ satisfies $(g,g)(i,j) = g \cdot (i,j) \cdot g^{-1} = (g(i), g(j))$ so that $(g,g)(P_i)$ is the pencil $P_{g(i)}$. This means that the diagonal group induces the full symmetric group on pencils while fixing e_G. In particular, we can find some $g' = (g,g) \in G \times G$ such that $\alpha := g'\alpha'$ fixes each pencil as a set, in addition to the vertex e_G. Let $x = (i,j)$ be an element of S_1. Then $\{x\} = P_i \cap P_j$ so that $\{\alpha(x)\} = \alpha(P_i) \cap \alpha(P_j) = \{x\}$. Hence α fixes all elements in $B_1(e_G)$ pointwise.

Note that $(1,2)$, $(1,3)$ and $(2,3)$ are pairwise joined to two elements in S_2, and no others, namely $x = (1,2,3)$ and $y = (1,3,2)$. Thus α fixes $\{x, y\}$ as a set and if $\imath$ denotes the inversion automorphism mentioned before, then either α or $\imath\alpha$ fixes all of $B_1(e_G) \cup \{(1,2,3)\}$ pointwise. By the following lemma either $\imath g'g''\alpha''$ or $g'g''\alpha''$ is the identity automorphism of Γ and so $\alpha'' = g''^{-1}g'^{-1}\imath$ or $\alpha'' = g''^{-1}g'^{-1}$ belongs to $(\mathrm{Sym}_n \times \mathrm{Sym}_n) \cdot C_2$. □

Lemma 5 *For $n \geq 3$ only the identity automorphism of $\mathrm{Sym}_n(T)$ fixes every vertex in $B_1(e_G) \cup \{(1,2,3)\}$.*

Proof: This is evident for $n = 3$. Suppose therefore that $n \geq 4$ and that α is an automorphism fixing every vertex in $B_1(e_G) \cup \{(1,2,3)\}$.

Then each double transposition in $(1^{n-4}2^2)^G$ is fixed by α as these elements have exactly two neighbours in S_1, with no two double transpositions having the same S_1-neighbours. The elements in $(1^{n-3}, 3^1)^G$ fall into pairs $[(i,j,k), (i,k,j)]$ of 3–cycles, each pairwise linked to the three fixed elements (i,j), (j,k) and (i,k) in S_1. Therefore α either fixes or interchanges the members in each pair. We show that α fixes these elements and hence is the identity on S_2.

Evidently $(1,2,3)$ and $(1,3,2)$ are both fixed. Hence look at the three pairs $[(1,4,2),(1,2,4)]$, $[(1,3,4),(1,4,3)]$ and $[(2,3,4),(2,4,3)]$. As can be calculated, the six 4-cycles in S_3 involving 1, 2, 3 and 4 are partitioned into two sets X, all connected to $(1,2,3)$, and Y, all connected to $(1,3,2)$. The sets X and Y are therefore fixed by α as sets. It turns out that $(1,4,2)$ is linked to two vertices in X while $(1,2,4)$ is linked to two vertices in Y. This means that $(1,4,2)$ and $(1,2,4)$ are each fixed by α. The same argument extends to all other 3–cycles. Hence $B_2(e_G)$ is fixed pointwise. For the remainder the argument becomes more homogeneous. Suppose that x and $y = \alpha(x)$ are in S_r with $r > 2$. By induction we can assume that α fixes all vertices in S_{r-1} and this means that x and y have the same neighbours $N(x) = N(y)$ in S_{r-1}. We claim that this forces $x = y$. The elements in $N(x)$ are obtained by 'splitting' any cycles appearing in x into two cycles in all possible ways, see (16). In particular, x and y have the same orbits on $\{1..n\}$ and if there are at least two orbits of length > 1 then $N(x) = N(y)$ forces $x = y$. In the remaining case x and y consist of a single cycle of length $\ell \geq 4$ with all other vertices fixed. It is easy to see that $\ell > 3$ and $N(x) = N(y)$ again forces $x = y$. □

6 Distance statistics in the transposition graph

Let S_i be the sphere of radius $i \leq n-1$ and centre e_G in the transposition Cayley graph $\mathrm{Sym}_n(T)$. Then S_i is a union of Sym_n-conjugacy classes and the parameters $a_i(y)$, $c_i(y)$ and $b_i(y)$ are constant on these classes, for all $1 \leq i \leq n-1$. It will be useful to set $s_i(n) := |S_i|$, and in more customary symbols, $c(n, n-i) := |S_i|$.

Then $c(n, n-i)$ is the number of permutations in Sym_n having $n-i$ cycles, for $1 \leq i \leq n-1$, and these are the *signless Stirling numbers of the first kind,* see for instance Chapter 1.3 in Stanley's book [20]. We have $c(n,n) = 1$, $c(n,n-1) = \binom{n}{2}$, $c(n,n-2) = 2\binom{n}{3} + 3\binom{n}{4}$, $c(n,n-3) = 3\binom{n}{4} + 20\binom{n}{5} + 15\binom{n}{6}$ and so on, up to $c(n,1) = (n-1)!$. The generating function of $c(n,m)$ satisfies

$$g(t) := \sum_{m=1}^{n} c(n,m)t^m = t(t+1)\cdots(t+n-1)$$

and from this we get the product form

$$\Pi_{\mathrm{Sym}_n(T)} = t^n g(t^{-1}) = (1+t)(1+2t)\cdots(1+(n-1)t)$$

for the Poincaré polynomial (5) of $\mathrm{Sym}_n(T)$. From the definition it is clear that $s_i(n)$ is a polynomial in n when i is fixed. The leading term counts the number of permutations of cycle type $1^{n-2i}\,2^i$ and so we note:

Lemma 6 *If i is fixed and $n \geq 2i$ then $s_i(n)$ is a polynomial in n of degree $2i$. Its leading term is the leading term of $\frac{1}{i!}\binom{n}{2}\binom{n-2}{2}\cdots\binom{n-2i+2}{2}$ and is equal to $\frac{1}{i!2^i}n^{2i}$.*

For $y \in S_i$ with cycle type $\mathrm{ct}(y) = 1^{h_1}2^{h_2}...\,n^{h_n}$ let as before $(1^{h_1}2^{h_2}...\,n^{h_n})^G = y^G$ be the conjugacy class of y. Then

$$|(1^{h_1}2^{h_2}...\,n^{h_n})^G| = \frac{n!}{1^{h_1}h_1!2^{h_2}h_2!\cdots n^{h_n}h_n!}\ ,$$

and

$$S_i = \bigcup_{h_1+h_2+\cdots+h_n=n-i} (1^{h_1}\,2^{h_2}\,...\,n^{h_n})^G\ , \tag{18}$$

see again Chapter 1.3 in [20]. Omitting cycle types of multiplicity 0 we therefore have $S_1 = (1^{n-2}\,2^1)^G$, $S_2 = (1^{n-3}\,3^1)^G \cup (1^{n-4}\,2^2)^G$ and so on. For small values of r one can compute $N(\mathrm{Sym}_n(T), r)$ easily from this information.

6.1 The value of $N(\mathrm{Sym}_n(T), r)$ for $r \leq 3$

As we have observed, $\mathrm{Sym}_n(T)$ is not distance-regular and as a consequence it is not straightforward to determine the value $N(\mathrm{Sym}_n(T), r)$ for general r. We begin to evaluate $N(\mathrm{Sym}_n(T), r)$ for $r \leq 3$ when closed formulae can be obtained.

Theorem 5 *For $n \geq 3$ we have*

$$N(\mathrm{Sym}_n(T), 1) = 3\,. \tag{19}$$

Proof: From Lemma 4 we have $\lambda(\mathrm{Sym}_n(T)) = 0$ since $a_1(z) = 0$ for $z \in S_1$ and moreover $c_2(y) = 3$ if y has cycle type $ct(y) = 1^{n-3}\,3^1$ and $c_2(y) = 2$ if y has cycle type $ct(y) = 1^{n-4}\,2^2$. Therefore, from (7) we have $\mu(\mathrm{Sym}_n(T)) = 3$ and by (10) we get (19). □

Theorem 6 *For $n \geq 5$ we have*

$$N(\mathrm{Sym}_n(T), 2) = N_2(\mathrm{Sym}_n(T), 2) = \frac{3}{2}(n+1)(n-2). \tag{20}$$

REMARK: From this result one can see that the bound in Theorem 2 is indeed very good: Working out the parameters for the transposition Cayley graph gives the bound $N(\mathrm{Sym}_n(T), 2) \geq N_2(\mathrm{Sym}_n(T), 2) \geq \frac{3}{2}(n+1)(n-2) - 1$ from Theorem 2.

Proof: By vertex transitivity it suffices to compute $|B_2 \cap B_2(y)|$ with $B_2 = B_2(e)$. This quantity depends only on the conjugacy class to which y belongs, this is a consequence of Proposition 2 and Theorem 4. Therefore we need to consider the number $N(y)$ of all vertices in B_2 which are at distance ≤ 2 from a given vertex $y \in S_i$ when i runs from 1 to 4. By (18) we have

$$S_4 = (1^{n-5}\,5^1)^G \cup (1^{n-6}\,2^1\,4^1)^G \cup (1^{n-6}\,3^2)^G \cup (1^{n-7}\,2^2\,3^1)^G \cup (1^{n-8}\,2^4)^G,$$
$$S_3 = (1^{n-4}\,4^1)^G \cup (1^{n-5}\,2^1\,3^1)^G \cup (1^{n-6}\,2^3)^G$$

and so on. The numbers $N(y)$ are presented in Table 1. The row index is the conjugacy class which contains y while the column index is the conjugacy classes contained in B_2. The value of $N(y)$ is worked out using (16).

N(y)	$(1^{n-3}\,3^1)^G$	$(1^{n-4}\,2^2)^G$	$(1^{n-2}\,2^1)^G$	$(1^n)^G$
$(1^{n-5}\,5^1)^G$	10	10	0	0
$(1^{n-6}\,2^1\,4^1)^G$	4	6	0	0
$(1^{n-6}\,3^2)^G$	2	9	0	0
$(1^{n-7}\,2^2\,3^1)^G$	1	7	0	0
$(1^{n-8}\,2^4)^G$	0	6	0	0
$(1^{n-4}\,4^1)^G$	4	2	6	0
$(1^{n-5}\,2^1\,3^1)^G$	1	3	4	0
$(1^{n-6}\,2^3)^G$	0	3	3	0
$(1^{n-3}\,3^1)^G$	$6(n-3)+2$	$3\binom{n-2}{2}$	3	1
$(1^{n-4}\,2^2)^G$	$4(n-2)$	$2\binom{n-2}{2}-1$	2	1
$(1^{n-2}\,2^1)^G$	$2(n-2)$	$\binom{n-2}{2}$	$\binom{n}{2}$	1

TABLE 1

When we consider the corresponding rows in the table we get $N_4(\mathrm{Sym}_n(T), 2) = 20$ when $n \geq 5$, $N_3(\mathrm{Sym}_n(T), 2) = 12$ when $n \geq 4$, $N_2(\mathrm{Sym}_n(T), 2) = \frac{3}{2}(n+1)(n-2)$ and $N_1(\mathrm{Sym}_n(T), 2) = (n-1)n$ for all $n \geq 3$. This proves the theorem due to (8). □

To estimate the number of vertices in $|B_r \cap B_r(y)|$ for an arbitrary y we consider the paths $t_1 t_2 \cdots t_{r^*}$ with $r^* \leq r$ starting at y and leading to a vertex $z = y t_1 t_2 \cdots t_{r^*}$ belonging to B_r. We say that this path has a *descent* at step $k < r^*$ if $y t_1 t_2 \cdots t_{k-1} \in S_s$ for some s while $y t_1 t_2 \cdots t_{k-1} t_k \in S_{s-1}$. The number of ways to continue the path at $y t_1 t_2 \cdots t_{k-1}$ by a descent is the downward degree $c(e_G, y t_1 t_2 \cdots t_{k-1})$ of (6) which by Lemma 4 is independent of n. Similarly, we say that the path has an *ascent* at step k if $y t_1 t_2 \cdots t_{k-1} \in S_s$ while $y t_1 t_2 \cdots t_{k-1} t_k \in S_{s+1}$. In this case the number of choices to continue the path at $y t_1 t_2 \cdots t_{k-1}$ by an ascent is the upward degree $b(e_G, y t_1 t_2 \cdots t_{k-1})$ which by Lemma 4 is of order n^2. Hence

Lemma 7 *The number of vertices $z = y t_1 t_2 \cdots t_{r^*}$ reachable from y on a path with a ascents is at most $k_y n^{2a}$ where k_y is some constant independent of n.*

Let $E_{i,i+1}$ be the set of edges joining a vertex in S_i to one in S_{i+1}. As $\mathrm{Sym}_n(T)$ is k-regular with $k = |S_1|$ and as $a_i(z) = 0$ for all $z \in S_i$, see Lemma 4, we have $|E_{i-1,i}| + |E_{i,i+1}| = k \cdot |S_i|$ and hence

$$|E_{r-1,r}| = k \cdot (|S_{r-1}| - |S_{r-2}| + |S_{r-3}| - ... + (-1)^{r-1}|S_0|) \tag{21}$$

for all r. For the transposition $y \in T$ let $E_{r-1,r}(y)$ be the set of all edges in $E_{r-1,r}$ of the form $\{z, zy\}$. (These are the edges in $E_{r-1,r}$ that are labelled by y.) Evidently all automorphisms of $\mathrm{Sym}_n(T)$ fixing e_G permute $E_{r-1,r}$ as a set and since the conjugation action is transitive on T every edge label must appear an equal number of times in each orbit. Hence

$$|E_{r-1,r}(y)| = |S_{r-1}| - |S_{r-2}| + |S_{r-3}| - ... + (-1)^{r-1}|S_0| \tag{22}$$

for all r. If $y = (j_1, j_2)$ then the end vertices $v^- \in S_{r-1}$ and $v^+ \in S_r$ of $\{v^-, v^+\} \in E_{r-1,r}(y)$ are composed of cycles in which j_1 and j_2 occur in different, respectively the same, cycle(s). Hence (22) gives the number of permutations in S_{r-1} with j_1, j_2 in different cycles and, at the same time, the number of permutations in S_r with j_1, j_2 in the same cycle.

Theorem 7 *Let $\Gamma = \mathrm{Sym}_n(T)$ be the transposition Cayley graph and suppose that $n \geq 4$. Then we have*

$$\begin{aligned}
(i) \quad N_1(\Gamma, 3) &= 2|S_0| + 2|S_2| \quad \text{and} \\
(ii) \quad N_2(\Gamma, 3) &= |S_0| + |S_1| + |S_2| + (n+2)(n-3) + \\
&+ 24\binom{n-3}{2} + 22\binom{n-3}{3} + 6\binom{n-3}{4}.
\end{aligned}$$

Furthermore we have $N(\Gamma, 3) = N_2(\Gamma, 3)$ for all $n \geq 16$.

Proof: We need to compute $|B_3 \cap B_3(y)|$ when $e = e_G$ and y have distance $d(e, y) \leq 6$ from each other. When $d(e, y) = 5$ or 6 then a path of length ≤ 3 from y to a vertex in B_3 can not have any ascents. Therefore the number of such vertices is independent of n by Lemma 7. (The same phenomenon can be seen in the upper part of Table 1.) When $d(e, y) = 3$ or 4 then the corresponding paths have at most one ascent so that $|B_3 \cap B_3(y)|$ is of order at most n^2. When $d(e, y) = 1$ or 2 then $|B_3 \cap B_3(y)|$ is of order at least n^4 as we will show now. It follows that the cases $3 \leq d(e, y) \leq 6$ can be ignored for large enough n, and a lower bound for n for this to be true will be given at the end of this proof.

(i) Finding $N_1(\mathrm{Sym}_n(T), 3)$: Let y be in S_1. Then by Lemma 1 we have $|B_3 \cap B_3(y)| = |B_2| + M(y)$ where $M(y)$ is the number of vertices $z \in S_3$ with $d(z, y) \leq 3$, and hence $d(z, y) = 2$. If $z = yt_1t_2 \in S_3$ with transpositions t_i then also $z^{-1} = t_2t_1y \in S_3$, see Theorem 4. Hence $M(y)$ is the number of all t_2t_1y belonging to S_3. As any element in S_2 is of the shape t_2t_1 for some t_1 and t_2 we see that $M(y) = |E_{2,3}(y)| = |S_2| - |S_1| + |S_0|$ by (22). Therefore

$$N_1(\mathrm{Sym}_n(T), 3) = |B_2| + |S_2| - |S_1| + |S_0| = 2|S_2| + 2|S_0|. \tag{23}$$

(ii) Finding $N_2(\mathrm{Sym}_n(T), 3)$: Let $y = y_1y_2$ be in S_2 with transpositions y_i. By Lemma 1 we have $|B_3 \cap B_3(y)| = |B_1| + M(y)$ where $M(y)$ is the number of vertices $z \in S_2 \cup S_3$ with $d(z, y) \leq 3$. As above, if $z = yt_1 \cdots t_{r^*} \in S_2 \cup S_3$ with $r^* \leq 3$, consider instead $z^{-1} = t_{r^*} \cdots t_1y_2y_1 \in S_2 \cup S_3$, that is, all paths from e_G of length ≤ 5 ending in y_2y_1 at a vertex in $S_2 \cup S_3$. Let Z be the set of all such vertices z^{-1}, in particular then $M(y) = |Z|$.

Let $u = t_2t_1 \in S_2$ be arbitrary. If $t_1 = y_1$ then $u = (t_2y_2)y_2y_1 \in Z \cap S_2$. Otherwise $uy_1 = (t_2t_1y_2)y_2y_1 \in Z \cap S_3$. Denote the vertices in Z of this kind by Z_0, in particular then $|Z_0| = |S_2|$.

Next let $e = \{v^+, v^-\} \in E_{2,3}(y_1)$ with $v^+ = v^-y_1 \in S_3$ and $v^- = v^+y_1 \in S_2$. Then v^- belongs to Z if and only if v^+y_2 belongs to S_2. Let the vertices of this type be denoted by Z_1. Thus $|Z_1|$ is the number of $u = v^+ \in S_3$ such that both uy_1 and uy_2 belong to S_2. When $y_1 = (1,2)$ and $y_2 = (2,3)$ then (16) implies that $|Z_1|$ is the number of elements in $(1,2,3)(4,5)^G$ or $(1,2,3,4)^G$ with 1, 2, 3 in the same cycle. This number is

$$|Z_1| = 2\binom{n-3}{2} + 6(n-3) = (n+2)(n-3)\ . \tag{24}$$

When $y_1 = (1,2)$ and $y_2 = (3,4)$ then $|Z_1|$ is the number of elements in $(1,2,3)(4,5)^G$, $(1,2,3,4)^G$ or $(1,2)(3,4)(5,6)^G$ in which 1, 2 and 3, 4 appear in the same cycle(s). This number is

$$|Z_1| = 4(n-4) + 6 + \binom{n-4}{2} = \binom{n}{2}\ . \tag{25}$$

Finally let $e = \{v^+, v^-\}$ be in $E_{3,4}(y_1)$ with $v^+ \in S_4$ and $v^- \in S_3$. Then v^- belongs to Z if and only if $u = v^+ \in S_4$ has the property that both uy_1 and uy_2 belong to S_3. Let Z_2 be the set of all such vertices v^-. When $y_1 = (1,2)$ and $y_2 = (2,3)$ then $|Z_2|$ is the number of elements in $(1,2,3)(4,5,6)^G$, $(1,2,3)(4,5)(6,7)^G$, $(1,2,3,4)(5,6)^G$ or $(1,2,3,4,5)^G$ with 1, 2, 3 in the same cycle. This number is

$$\begin{aligned} |Z_2| &= 4\binom{n-3}{3} + \binom{n-3}{2}\binom{n-5}{2} \\ &+ 3!(n-3)\binom{n-4}{2} + 4!\binom{n-3}{2} \\ &= 24\binom{n-3}{2} + 22\binom{n-3}{3} + 6\binom{n-3}{4}. \end{aligned} \tag{26}$$

(Note, the term $\binom{n-3}{2}\binom{n-5}{2}$ accounts for the two choices of a 3-cycle on $\{1,2,3\}$ while avoiding duplication in the choice of two 2-cycles from the remaining $n-3$ and $n-5$ vertices, respectively.)

When $y_1 = (1,2)$ and $y_2 = (3,4)$ then $|Z_2|$ is the number of elements in $(1,2,3)(4,5,6)^G$, $(1,2,3)(4,5)(6,7)^G$, $(1,2,3,4,5)^G$, $(1,2,3,4)(5,6)^G$ or $(1,2)(3,4)(5,6)(7,8)^G$ in which 1, 2, and 3, 4 appear in the same cycle(s). This number is

$$\begin{aligned} |Z_2| &= 4(n-2)(n-5) + \left[4(n-4)\binom{n-5}{2} + 2\binom{n-4}{3}\right] + \\ &+ 4!(n-4) + \left[6\binom{n-4}{2} + 6\binom{n-4}{2} + 6\binom{n-4}{2}\right] + \\ &+ \frac{1}{2}\binom{n-4}{2}\binom{n-6}{2} \\ &= 24(n-4) + (n-5)(13n-44) + 14\binom{n-4}{3} + 3\binom{n-4}{4}. \end{aligned} \tag{27}$$

It is clear that $Z = Z_0 \cup Z_1 \cup Z_2$ is a disjoint union. Comparing (24)+(26) to (25)+(27) one can check that the first expression is bigger than the second for all $n \geq 4$. Hence $|B_3 \cap B_3(y)|$ takes

its maximum when $y \in (1,2,3)^G$ for all $n \geq 4$. Therefore

$$\begin{aligned} N_2(\mathrm{Sym}_n(T),\,3) &= |S_0| + |S_1| + |S_2| + (n+2)(n-3) + \\ &+ 24\binom{n-3}{2} + 22\binom{n-3}{3} + 6\binom{n-3}{4} \end{aligned} \tag{28}$$

and this complete the second part of the theorem.

We now return to the comment at the beginning of this proof. Comparing (23) to (28) shows that $N_2(\Gamma,\,3) > N_1(\Gamma,\,3)$ for all $n \geq 4$. When y belongs to S_3 or S_4 a very rough upper bound for $|B_3 \cap B_3(y)|$ can be obtained by following through the argument in Lemma 7. By Lemma 4 the downward degree for a vertex in $S_j(e)$ is at most $\binom{j+1}{2}$ while the upward degree is at most $\binom{n}{2} - j$. By considering the possible ascent-descent combinations of a path from y to a vertex in B_3 we can work out that $N_4(\Gamma,3) \leq -455 + 155\binom{n}{2}$ and $N_3(\Gamma,3) \leq -132 + 65\binom{n}{2}$. Using the same arguments we can bound $N_6(\Gamma,3) \leq 1575$ and $N_5(\Gamma,3) \leq 525$. Evaluating these inequalities it can be seen that $N_2(\Gamma,\,3) > N_j(\Gamma,\,3)$ for $j = 3,\,4,\,5,\,6$ from $n \geq 16$ onwards. We note that a better lower bound $n \geq 10$ can be obtained by a more careful albeit tedious count of the possible paths. This completes the proof. □

6.2 The asymptotic behaviour of $N(\mathrm{Sym}_n(T), r)$

The main work in this section will be to find $N_1(\mathrm{Sym}_n(T), r)$ and $N_2(\mathrm{Sym}_n(T), r)$ for arbitrary r and sufficiently large n. It will turn out that this determines $N(\mathrm{Sym}_n(T), r)$ in general. Let $b(n,r)$ denote the cardinality of the ball of radius r in $\mathrm{Sym}_n(T)$, thus

$$b(n,r) = |B_r| = \sum_{0 \leq i \leq n} s_i(n) = \sum_{0 \leq i \leq n} c(n, n-i)$$

in terms of the signless Stirling numbers of the first kind.

First we consider $N_2(\mathrm{Sym}_n(T), r)$. By Lemma 1 we have $N_2(\mathrm{Sym}_n(T), r) = b(n, r-2) + |(S_r \cup S_{r-1}) \cap B_r(y^*)|$ where y^* suitably is a 3-cycle or a double transposition. We set $A := |(S_r \cup S_{r-1}) \cap B_r(y^*)|$ and let $y^* = y_1 y_2$ with two transpositions y_i. We now need to find the number A of all z in $S_r \cup S_{r-1}$ which can be reached on a path $t_1 t_2 ... t_{r^*}$ of length $r^* \leq r$ starting from y^*. This means that $z = y_1 y_2 t_1 t_2 ... t_{r^*}$ and applying the inversion automorphism, see Theorem 4, we obtain the element $z^{-1} = t_{r^*} \cdots t_2 t_1 y_2 y_1$ in $S_r \cup S_{r-1}$. This represents a path from $e = e_G$ in which y_2 and y_1 are the last edges. If Z denotes the set of all such elements z^{-1} then $|Z| = A$.

Let u be in S_{r-1} and suppose that $u = t_{r-1} \cdots t_2 t_1$ is the product of suitable transpositions t_i. If $t_1 = y_1$ then $u = (t_{r-1} \cdots t_2 y_2) y_2 y_1$ so that $u \in Z$. Otherwise $u y_1 = (t_{r-1} \cdots t_2 t_1 y_2) y_2 y_1$ belongs to Z. Thus every u in S_{r-1} gives rise to one element in Z. The set of elements of this type is denoted by Z_0, and in particular $|Z_0| = |S_{r-1}|$.

The next type of vertices in Z are of the shape $z^{-1} = t_{r^*} \cdots t_2 t_1 y_2 y_1$ where both z^{-1} and $t_{r^*} \cdots t_2 t_1$ belong to S_{r-1} while $t_{r^*} \cdots t_2 t_1 y_2$ belongs S_r. This type will be called Z_1, evidently this set is disjoint from Z_0.

The remaining vertices in Z are of the shape $z^{-1} = t_{r^*} \cdots t_2 t_1 y_2 y_1$ where both z^{-1} and $t_{r^*} \cdots t_2 t_1$ belong to S_r while $t_{r^*} \cdots t_2 t_1 y_2$ belongs S_{r+1}. These are the vertices of type Z_2 and it follows that $Z = Z_0 \cup Z_1 \cup Z_2$ is a disjoint union. Therefore

$$N_2(\mathrm{Sym}_n(T), r) = b(n, r-1) + \max(|Z_1| + |Z_2| \ : \ y^* \in S_2) \tag{29}$$

where Z_1 and Z_2 depend on the choice of y^* as either a 3-cycle or a double transposition. We can now prove the following theorem:

Theorem 8 *Let $\Gamma = \mathrm{Sym}_n(T)$ be the transposition Cayley graph. Suppose that $r \geq 2$ and that y is a transposition.*

(i) For $n - 1 \geq r$ we have

$$\begin{aligned} N_1(\Gamma,\, r) &= b(n, r-1) + |E_{r-1,r}(y)| = \\ &= 2 \cdot \left(|S_{r-1}| + |S_{r-3}| + |S_{r-5}| + \cdots \right) . \end{aligned} \tag{30}$$

(ii) For n sufficiently large we have

$$N_2(\Gamma,\, r) \ > \ N_1(\Gamma,\, r) \ . \tag{31}$$

Proof: (i) Let y be in S_1. By Lemma 1 we have $N_1(\Gamma, r) = |B_r \cap B_r(y)| = |B_{r-1}| + |S_r \cap B_r(y)|$. Hence we need to find the number $M(y) = |S_r \cap B_r(y)|$ of all $z \in S_r$ which can be reached on a path of length $\leq r$ from y. Such a path is necessarily of the shape $z = y t_1 t_2 \cdots t_{r-1}$, consisting of $r-1$ transpositions t_i. Applying the inversion automorphism, as above, we obtain a path $z^{-1} = t_{r-1} \cdots t_2 t_1 y$ starting from e to $z^{-1} \in S_r$ with y as last edge. Evidently any vertex in S_{r-1} can be reached by a suitable choice of $t_{r-1} \cdots t_2 t_1$ and therefore $M(y) = |E_{r-1,r}(y)|$. The result now follows from (22).

(ii) Let Z_1 and Z_2 have the same meaning as in (29). By the first part of this theorem and the equation (29) it will be sufficient to show that $|Z_2| \geq |E_{r-1,r}(y)|$ for all large enough n. We evaluate $|Z_2|$ for $y^* = y_1 y_2$ with $y_1 = (1,2)$ and $y_2 = (2,n)$. The vertices in Z_2 are in one-to-one correspondence with vertices $u \in S_{r+1}$ such that uy_1 and $uy_2 \in S_r$. By the basic property (16) this is the case if and only if $1, 2, n$ belong to the same cycle of u. Let U be the set of such elements, $|U| = |Z_2|$. To count $|U|$ consider elements u' in the symmetric group G' on $\{1, 2, ..., n-1\}$ which have the following properties: (i) both 1 and 2 are in the same cycle of u', and (ii) u' has $(n-1) - r$ cycles. Thus u' is in $S_r \cap G'$ and it follows from (22) that the total number of such elements u' is

$$\begin{aligned} a: \ &= \ | \, E_{r-1,r}(y_1) \ \cap \ \{ \, \{x,\, x^*\} \ : \ x,\, x^* \in G' \} \, | \\ &= \ |S_{r-1} \cap G'| - |S_{r-2} \cap G'| + |S_{r-3} \cap G'| - ... + (-1)^{r-1} |S_0 \cap G'| \ . \end{aligned}$$

By Lemma 6 it follows that the coefficient of the leading power of n in $|S_{r-1}|$ and in $|S_{r-1} \cap G'|$ is the same. Therefore

$$a = |E_{r-1,r}(y_1) \cap \{ \{x, x^*\} \ : \ x, x^* \in G' \} \, | = |S_{r-1}| + f(n)$$

for a polynomial $f(n)$ of degree $\leq 2(r-1)-1$. Any u' of the kind just considered is a permutation on $\{1,2,...n-1\}$ in which 1 and 2 appear in the same cycle, say of length $c_{u'} \geq 2$. Now we may insert n into that cycle, in $c_{u'}$ distinct ways, to get $c_{u'}$ different elements in U. Thus $|Z_2| = |U| \geq 2a$ and therefore

$$|Z_2| \geq 2 \cdot |S_{r-1}| + 2f(n)\,. \tag{32}$$

Since the leading term of $|E_{r-1,r}(y)|$ is $|S_{r-1}|$ by (22) it follows that $|Z_2| \geq |E_{r-1,r}(y)|$ for all large enough n. □

In the expression $N_2(\mathrm{Sym}_n(T), r) = b(n, r-1) + \max(|Z_1| + |Z_2| \ : \ y^* \in S_2)$ stated in (29) the term $|Z_1| + |Z_2|$ depends on the choice of y^*. We therefore turn to evaluating $|Z_1| + |Z_2|$ for the two possible choices $y^* = (1,2,3)$ and $y^* = (1,2)(3,4)$.

This leads us to the following definition. Let $c_{3^1}(n, n-i)$ be the number of vertices in S_i in which the letters $1,2,3$ appear in a single cycle, and let $c_{2^2}(n, n-i)$ be the number of vertices in S_i in which the letters $1,2$ and $3,4$ appear in the same cycle or cycles. For instance,

$$\begin{aligned} c_{3^1}(n,n) &= c_{3^1}(n,n-1) = 0, \quad c_{3^1}(n,n-2) = 2, \\ c_{3^1}(n,n-3) &= (n+2)(n-3), \qquad \text{and} \\ c_{3^1}(n,n-4) &= 24\binom{n-3}{2} + 22\binom{n-3}{3} + 6\binom{n-3}{4}\,. \end{aligned} \tag{33}$$

Similarly we have

$$\begin{aligned} c_{2^2}(n,n) &= c_{2^2}(n,n-1) = 0, \quad c_{2^2}(n,n-2) = 1, \\ c_{2^2}(n,n-3) &= \binom{n}{2}, \qquad \text{and} \\ c_{2^2}(n,n-4) &= 24(n-4) + (n-5)(13n-44) \\ &+ 14\binom{n-4}{3} + 3\binom{n-4}{4}. \end{aligned} \tag{34}$$

For this see again (24), (25), (26) and (27). As we have already observed in the proof of the last theorem, the general rule (16) implies that Z_1 and Z_2 in (29) satisfy

$$|Z_1| + |Z_2| = c_{3^1}(n, n-r) + c_{3^1}(n, n-(r+1)) \tag{35}$$

when y^* is a 3-cycle and

$$|Z_1| + |Z_2| = c_{2^2}(n, n-r) + c_{2^2}(n, n-(r+1)) \tag{36}$$

when y^* is a double transposition. We obtain an estimate for $c_{3^1}(n, n-r)$ and $c_{2^2}(n, n-r)$ as follows:

Lemma 8 *For fixed $i \geq 2$ and n sufficiently large we have*

$$c_{3^1}(n, n-i) = \frac{2}{(i-2)!}\binom{n-3}{2}\binom{n-5}{2}\cdots\binom{n+3-2i}{2} + f_1 \tag{37}$$

and

$$c_{2^2}(n, n-i) = \frac{1}{(i-2)!}\binom{n-4}{2}\binom{n-6}{2}\cdots\binom{n+2-2i}{2} + f_2 \tag{38}$$

where the f_i *are polynomials in* n *of degree* $< d_i = 2(i-2)$. *In particular,* $c_{3^1}(n, n-i)$ *and* $c_{2^2}(n, n-i)$ *are polynomials of degree* d_i *and* $c_{3^1}(n, n-i) = 2\,c_{2^2}(n, n-i) + f_3$ *with a polynomial* f_3 *of degree* $< d_i$.

Proof: Let $C_{3^1}(n, n-i) \subseteq S_i$ and $C_{2^2}(n, n-i) \subseteq S_i$ be the sets counted by $c_{3^1}(n, n-i)$ and $c_{2^2}(n, n-i)$ respectively. For $i = 2$, when $0! = 1$, we see that $C_{3^1}(n, n-2)$ consists of the two 3-cycles $(1, 2, 3)$ and $(1, 3, 2)$ while $C_{2^2}(n, n-2)$ consists of the single double transposition $(1, 2)(3, 4)$ only. This established the base of induction and accounts for the factor 2 throughout. We will prove the statement (37) concerning $c_{3^1}(n, n-i)$, the corresponding statement (38) for $c_{2^2}(n, n-i)$ follows in exactly the same way.

If $g \in \mathrm{Sym}_n$ let $\mathrm{supp}(g)$ be its *support,* that is all symbols moved by g. The cardinality of the support of any $g \in C_{3^1}(n, n-i)$ is at most $3+2(i-2)$. Let $C_0^i := C_{3^1}(n, n-i) \cap (1^{n-2i+1}2^{i-2}3^1)^G$ and let $C_1^i = C_{3^1}(n, n-i) \setminus C_0^i$. Then

$$|C_0^i| = \frac{2}{(i-2)!}\binom{n-3}{2}\binom{n-5}{2}\cdots\binom{n+3-2i}{2}$$

and by induction we assume that $|C_1^i| = f_1$ has degree $< d_i$. Since

$$|C_0^{i+1}| = \frac{2}{(i-1)!}\binom{n-3}{2}\binom{n-5}{2}\cdots\binom{n+3-2i}{2}\cdot\binom{n+1-2i}{2}$$

it remains to show that the number of elements in C_1^{i+1} is a polynomial of degree at most $d_i + 1$.

By considering cycle types it is easy to see that any vertex in C_1^{i+1} has at least one neighbour in C_0^i or in C_1^i. In the first case, if $g = u \cdot (j_1, j_2)$ with $u \in C_0^i$ then at least one of $\{j_1, j_2\}$ must be in the support of g as otherwise $g \in C_0^{i+1}$. The number of such elements g therefore is polynomial of degree at most $d_i + 1$. The number of vertices of the second kind is clearly at most $f_1\binom{n}{2}$, again of degree at most $d_i + 1$. Hence $c_{3^1}(n, n-i)$ has the required expression. In the case of $c_{2^2}(n, n-i)$ the same arguments apply. □

Theorem 9 *Let* $\Gamma = \mathrm{Sym}_n(T)$ *be the transposition Cayley graph and suppose that* $r \geq 1$. *Then for all sufficiently large* n *we have*

$$\begin{aligned} N(\Gamma, r) &= N_2(\Gamma, r) \\ &= b(n, r-1) + c_{3^1}(n, n-r) + c_{3^1}(n, n-(r+1))\,. \end{aligned} \tag{39}$$

REMARK: For $r \leq 3$ we already have computed the value of $N(\mathrm{Sym}_n(T), r)$ in Theorems 5, 6 and 7 when $N(\mathrm{Sym}_n(T), r)$ indeed agrees with (39). In these theorems the lower bound on n was explicit and hence better than the condition here. Nevertheless, analysing the arguments here it is likely that the bound $n > 3r$ is sufficient.

Proof: We can assume that $r > 3$. By (29), (35) and Lemma 8 it follow that $N_2(\Gamma, r) = b(n, r-1) + c_{3^1}(n, n-r) + c_{3^1}(n, n-(r+1))$ and this establishes the second equation. By

Theorem 8 we know that $N_2(\mathrm{Sym}_n(T), r) > N_1(\mathrm{Sym}_n(T), r)$ and both terms are polynomial of degree $2(r-1)$. It remains to show that $N_s(\mathrm{Sym}_n(T), r)$ is polynomial of degree $< 2(r-1)$ for $2 < s$. For $y \in S_s$ consider all paths $z = yt_1 \cdots t_{r^*}$ of length $r^* \leq r$ to a vertex $z \in S_{s^*}$ with $s^* \leq r$. If a and b are the number of ascents and descents then $a+b = r^* \leq r$ and $a-b = s^*-s$. From this it follows that $a < r-1$ and the required fact now follows from Lemma 7. □

References

[1] *Topics in Algebraic Graph Theory*, Encyclopedia of Mathematics and its Applications, Volume 132, edited by L. W. Beineke and R.J. Wilson, Cambridge University Press, 2004.

[2] L. C. Grove and C.T. Benson, Finite Reflection Groups. Second edition. Graduate Texts in Mathematics, Volume 99. Springer-Verlag, New York, 1985.

[3] A. Björner and F. Brenti, *Combinatorics of Coxeter Groups,* Springer Verlag, Heidelberg, New York, 2005.

[4] A. E. Brouwer, A. M. Cohen and A. Neumaier, *Distance–regular graphs*, Springer-Verlag, Berlin, Heidelberg, 1989.

[5] P. Diakonis, Group Representations in Probability and Statistics, Institute of Mathematical Statistics, Volume 11, Hayward, California, 1985.

[6] J. Denes, Representation of a permutation as the product of a minimal number of transpositions, and its connection with the theory of graphs, *Publ. Math. Institute Hung. Acad. Sci.* **4** (1959) 63–70.

[7] D. J. Futuyma, Evolutionary Biology, 3rd edition, Sinauer Associates, 1998.

[8] L. Heydemann, *Cayley graphs,* In: G. Hahn, G. Sabidussi (eds.), *Graph Symmetry: Algebraic Methods and Applications,* Kluwer, Amsterdam, 1997.

[9] S. Lakshmivarahan, J. Jwo, and S. K. Dhall, Symmetry in interconnection networks based on Cayley graphs of permutation group: a survey, *Parallel Comput.* **19** (1993) 361–407.

[10] F.R.C. Chung and R.P. Langlands, A combinatorial Laplacian with vertex weights, *Journal of Combin. Theory, Ser. A* **75** (1996) 316–327.

[11] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions and reversals, (in Russian), *Dokl. Acad. Nauk* **163** 4 (1965) 845–848; English translation, *Sov. Phys.-Dokl.* **10** 8 (1966) 707–710.

[12] V. I. Levenshtein, Reconstructing objects from a minimal number of distorted patterns, (in Russian), *Dokl. Acad. Nauk* **354** (1997) 593–596; English translation, *Doklady Mathematics* **55** (1997) 417–420.

[13] V. I. Levenshtein, Efficient reconstruction of sequences, *IEEE Trans. Inform. Theory* **47** 1 (2001) 2–22.

[14] V. I. Levenshtein, Efficient reconstruction of sequences from their subsequences or supersequences, *Journal of Combin. Theory, Ser. A* **93** 2 (2001) 310–332.

[15] P. Maynard and J. Siemons, Efficient reconstruction of partitions, *Discrete Mathematics* **293** (2005) 205–211.

[16] B. Nobel and J.W. Daniel, Applied Linear Algebra, second edition, Prentice Hall, New Jersey, 1977.

[17] P. A. Pevzner, *Computational molecular biology: an algorithmic approach*, The MIT Press, Cambridge, MA, 2000.

[18] O. Pretzel and J. Siemons, Reconstruction of partitions. Electron. J. Combin. 11 (2004/06), no. 2, Note 5, 6 pp. (electronic).

[19] D. Sankoff and N. El–Mabrouk, Genome rearrangement, In: *Current topics in computational molecular biology*, Eds.: T. Jiang, T. Smith, Y. Xu and M.Q. Zhang, MIT Press, 2002.

[20] R. Stanley, *Enumerative Combinatorics, Vols I* and *II,* Cambridge University Press, 1997 and 1999.

[21] J.H. van Lint and R.M. Wilson, *A Course in Combinatorics,* Cambridge University Press, 1992.

[22] V. G. Vizing, On estimates of the chromatic class of a p-graph, *Discretnyi Analiz,* **3** (1964) 25–30 (in Russian).